Article:

Strachan, James, Kirkham, Alexander, Manssuer, Luis et al. (1 more author) (2016) Incidental learning of trust : Examining the role of emotion and visuomotor fluency. Journal of Experimental Psychology: Learning, Memory, and Cognition. pp. 1759-1773. ISSN 1939-1285

https://doi.org/10.1037/xlm0000270

Incidental learning of trust: Examining the role of emotion and visuomotor fluency.

James W. A. Strachan[1], Alexander J. Kirkham[1], Luis R. Manssuer[2] and Steven P. Tipper[1]

[1] University of York, York, England

[2] Bangor University, Bangor, Wales

Corresponding author:

J. W. A. Strachan,

Department of Psychology, University of York,

York, North Yorkshire, YO10 5DD.

Contact: js756@york.ac.uk

Abstract

Eye gaze is a powerful directional cue that automatically evokes joint attention states. Even when faces are ignored there is incidental learning of the reliability of the gaze cueing of another person, such that people who look away from targets are judged less trustworthy. In a series of experiments we demonstrate further properties of the incidental learning of trust from gaze direction. First, the emotion of the face, whether neutral or smiling, influences the pattern of trust learning. Second, the effect is specific to judgements of trust; reliability of gaze direction does not influence other emotional judgements of a person, such as liking. And third, visuomotor fluency is not sufficient for learning of trust, regardless of whether the face serves as a target or distractor. Taken together, incidental learning of trust is influenced by facial emotion, it is a specific effect that does not generalize to other emotional assessments, and it is not determined solely by processing fluency.

Keywords: gaze-cueing; task-switching; expression; trait inference; visuomotor fluency; emotion; trustworthiness; likeability

There is a scene in the 2006 James Bond film *Casino Royale* where Daniel Craig's Bond is struggling with another man over a knife. During the fight, Bond hesitates and stares over the other man's shoulder, distracting his opponent temporarily and granting him the upper hand. For most people, fights to the death are thankfully uncommon, but this misdirection of one's attention is a common trope that we all easily recognise. Redirection of our attention based on somebody else's can be perfectly innocent, as when someone genuinely spots something that demands their and perhaps your own attention, or it can be the result of a calculated deception, as in the case of sleight of hand magic tricks. Interpreting such events in terms of their social implications therefore reflects a key element of social cognition.

Hence eye gaze is a powerful communicative tool. It can be used to reflexively redirect another's attention towards or away from a particular object or location (Driver et al., 1999; Friesen & Kingstone, 1998; Frischen & Tipper, 2004). Due to its dual ability to either facilitate attentional processing or misdirect it and incur a cost to processing fluency, people can also infer higher order information from this; objects that are looked at tend to be liked more (Bayliss, Paul, Cannon and Tipper, 2006; Capozzi, Bayliss, Elena and Becchio, 2015; Manera, Elena, Bayliss and Becchio, 2014; Ulloa, Marchetti, Taffou and George, 2015) and those who correctly cue an object location are chosen as more trustworthy than those who mislead (Bayliss & Tipper, 2006; Bayliss, Griffiths & Tipper, 2009; Manssuer, Pawling, Hayes & Tipper, 2015; Rogers, Bayliss et al., 2014).

In the initial investigation of incidental learning of trust from gaze behaviour, Bayliss and Tipper (2006) used a gaze-cueing paradigm with a group of paired faces. One face of each pair would always look towards a target object that participants had to identify as either a kitchen or garage item (valid cue), while the other would always look away from the target (invalid cue). At the end of the experiment, participants were shown each face in the pair and asked to select which they thought was more trustworthy. Despite having been told that the gaze behaviour of the face was task-irrelevant and that they should ignore the face throughout the experiment, participants chose the valid cueing faces over invalid cueing faces as more trustworthy.

This effect has been shown using this 2-alternative forced choice (2AFC) rating procedure. However, this does not explain how these changes come about: whether this effect is driven by an increase in trustworthiness for valid faces, a decrease for invalid faces, or a bidirectional mix of the two. To further investigate the specific nature of changes in trust ratings two scalar ratings of trustworthiness will be employed in the current studies, one at the beginning and one at the end of the experiment, to track changes in trustworthiness for both valid and invalid faces (c.f., Manssuer, Roberts & Tipper, 2015). This more sensitive measure provides the ideal approach to further investigate key boundary conditions for the understanding of the processes mediating incidental learning of trust.

One outstanding issue where this new measure may be benefial concerns the role of facial emotion. Bayliss, Griffiths and Tipper (2009) found that gaze-contingent trust effects appear to rely on a positive social context, as they found no trust effects when the faces expressed anger and a reliable effect only when the faces smiled. However, the neutral expression condition was somewhat ambiguous, as participants were only slightly more likely to select the valid face as the more trustworthy of a matched pair in a 2AFC paradigm. This previous work using forced choice between valid and invalid cueing individuals creates a somewhat blunt measure; we can see that valid faces are preferred over invalid, but we do not know if this is because valid faces become more trustworthy, invalid become less, or some combination of the two. We also do not know how exactly the emotional expression of a face might change the *pattern* of results using forced-choice measures of trustworthiness. It still could be the case that neutral faces are not sufficient to elicit learning of trust – there is a wealth of evidence that smiling faces are treated differently from neutral faces in various social interactions, both when measured by trustworthiness judgements (Hehman, Flake & Freeman, 2015) and by more implicit measures (Wang & Hamilton, 2014) – or it may be that we can detect trust learning with neutral faces using this new, more sensitive measure.

Therefore we examine the role of emotion in the incidental learning of trust in conditions where faces express neutral emotions (Experiment 1) and when they smile (Experiment 2). We will be able to unequivocally identify whether incidental learning of trust from gaze cueing can be detected when faces express neutral emotions. Additionally, and more importantly, we can assess whether the pattern of learning (whether valid faces increase in trust and invalid faces decline in trustworthiness, or whether the effects are uni-directional) is the same for both neutral and positive emotions.

The further issue we investigate is whether incidental learning of eye-gaze patterns is specific to judgements of trust, or generalizes to other emotional assessments, such as liking of a person. One might assume that trust and liking will be closely related: if we trust someone, we are more likely to like them. Indeed, the two are often conflated as aspects of warmth in dual-dimension theories of social cognition (e.g. Fiske, Cuddy & Glick, 2007). However, subtle behaviours that can be used to deceive others, such as gaze shifts, could have quite specific effects on trust. For example, whether to invest money with another person is influenced by incidental learning of patterns of gaze shifts, as is the decision to be altruistic while computing the likelihood that such an act will be reciprocated in the future

(Rogers, Bayliss et al., 2014). Such decisions might not be affected by general feelings of liking, for example we may trust a lawyer to do their utmost to preserve our freedom, but we may not like them on a personal level; the two feelings are distinct, and can be separated. To our knowledge, there is little previous work directly addressing the question of whether trust and liking are functionally similar in this way, so in Experiment 3 we replace the trustworthiness ratings of Experiment 1 with likeability ratings, to see if this incidental learning is specific to trust or if there is a broader affective spillover into other social judgements.

The final issue concerns the role of visuomotor fluency in the learning of trust. In the gaze cueing procedure there are two aspects that might mediate the learning of trust. One of these is the behaviour of another person. As noted, gaze can be used to help or deceive another person. That is, looking towards interesting and desirable objects to facilitate a conspecific's behaviour, or looking away from desirable objects to mislead. The second aspect is the visuomotor fluency experienced during gaze cueing. That is, responses are faster on valid trials where gaze orients a person's attention to the location where a target will appear. Previous work has shown that facilitating both perceptual (e.g., Reber, Winkielman & Schwartz, 1998) and motor (e.g., Hayes, Paul, Beuger & Tipper, 2008) performance increases preference and liking of images and objects.

Therefore we investigate the incidental learning of trust in a task where increased processing fluency is associated with some faces and impaired processing fluency is associated with other faces in a similar way to gaze cueing studies. However, there are no face behaviours, such as gaze shifts, that might be associated with deception. To this end we develop two new task switching procedures in Experiments 4 and 5, and consistently associate some face identities with fluent, fast and accurate processing and other faces with impaired slow and error prone processing. In Experiment 4 we develop a task-switching procedure where the faces are now the targets of participants' decisions, and in Experiment 5 we use a design where the faces remain background distractors. If visuomotor fluency is the key driver for the learning of trust, then effects should be detected in one or both of these new tasks. On the other hand, if cues to deception such as eye-gaze are necessary, then no learning of trust will be detected.

To briefly preview our findings: We find that face emotion does influence the learning of trust from gaze behaviour, but in a selective manner. That is, when faces express neutral emotions, there is a decline in trust for invalid faces that look away from targets. No such effect is observed with valid faces, except for when these faces are seen smiling. When examining the generalizability of the trust effect, somewhat counter-intuitively, we find no effects when assessing liking of another person. And finally, in both task switching procedures, where the faces were targets, and where they were distractors, there were no changes in trust assessments, suggesting that changes in visuomotor fluency are not sufficient to generate learning of trust.

Experiment 1

We re-examine whether incidental learning of trust can be detected with faces expressing neutral emotion, and if such an effect is detected, what pattern of changes in trust are revealed?

Methods

Participants

A total of 24 participants (18 female) volunteered for the study in return for payment or course credit. All were students of the University of York, and had a mean age of 19.96

years. All participants in all experiments described in this study provided written consent and the research was given ethical approval by the Departmental Ethics Committee of the University of York Psychology Department.

Stimuli

Target stimuli for the object classification task were the kitchen and garage object images used in Bayliss and Tipper (2006). There were 13 unique objects in each group (kitchen/garage) that appeared in both left and right orientations. All of the stimuli were coloured in blue. In total, there were 52 individual images used in the experiment. Face stimuli were taken from the Karolinska Directed Emotional Faces (KDEF) stimulus set (Lundqvist, Flykt & Öhman, 1998), and included sixteen images; eight male and eight female. These faces were initially selected by eye from a figure in the Supplementary Material of Oosterhof and Todorov (2008), in which the faces from this set are plotted along six judgement dimensions. The faces used were all taken from the centre (1SD from the intercection of all six dimensions) of this plot, so the faces used in our experiments were, compared with the rest of the KDEF set, as close to neutral trait judgements as possible.[1] These faces were split into two groups, which would appear as either valid or invalid cues in the experiment (counterbalanced across participant). The eyes of each face were manipulated using Adobe Photoshop CS6 to generate faces where the eye gaze was either straight ahead, left or right. Unaltered images were used for the trustworthiness rating sections.

The study was run on an Intel Core i5 PC with a 21.5” monitor. The experiment was presented using E-Prime 2.0 software with a white background throughout and the resolution set to 1024x768 pixels. Participants were sat approximately 60cm from the display, and during trustworthiness ratings the face stimuli had a visual angle of 19.29°

horizontally and 20.97° vertically, while during gaze-cueing the face stimuli had a visual angle of 13.36˚ horizontally and 14.93˚ vertically.

Design and Procedure

Participants were told that they would be asked to perform an object categorisation task on images of objects that appeared on the left or right side of the screen, and to respond with whether these were garage or kitchen objects. They were also told that the central face images were irrelevant and to be ignored. Before the experiment participants were allowed to study printed versions of the kitchen/garage images, in order to familiarise themselves. This was done firstly to ensure that participants had the knowledge of what each object was, and secondly to make sure that early responses from the first trial block were not confounded by uncertainty as to the object categories of the targets.

[Figure 1 approximately here]

Each trial began with a 600ms fixation cross in the centre of the screen, which was then replaced by a face showing a direct gaze for 1,500ms. The face then shifted gaze either to the left or right for 500ms before the target stimulus appeared on either the same (valid) or opposite (invalid; see Figure 1a) side of the gaze direction. The target stimulus remained either until the participant's response was logged or until 2,500ms had passed, following which participants received feedback from an error tone that would sound if an incorrect response was logged. The face then shifted back to direct gaze for a further 1,000ms. A blank screen followed for 500ms before the next trial began. The trial structure is shown in Figure 1b.

The object categorisation responses were the H key and the Space bar of a keyboard, chosen because the H key appears directly above the Space bar on QWERTY keyboards and this direction was orthogonal to the possible location of the target. Participants were instructed to respond with their index finger on the H key and thumb on the Space bar. For half of the participants, H represented Kitchen objects, while for the other half it represented Garage objects.

In total there were five blocks of 32 trials each, and each face appeared twice in each block, once gazing left and once right (ten times in total across the experiment; five left, five right). The order of faces was randomised, as was the order of target objects, the side that the target appeared, and the order of valid and invalid trials.

At the beginning and the end of the experiment, participants rated the original unmanipulated face images used to generate the gaze cueing stimuli. Participants were shown a calibration slide where they clicked in the centre to start, and then the face images were presented for 1,000ms. Participants were then instructed to click along an uninterrupted scale at a point that conformed to how trustworthy they felt the person was. The scale recorded response clicks between -100 and +100, calculated by the distance from the centre of the scale − responses to the left of the centre of the scale were coded as negative, while those to the right were coded as positive (these were indicated on the screen with a − and + sign at either end of the scale). Identities were presented in a randomised order.

After the experiment, we asked all participants what they thought the nature of the experiment had been and if they had picked up on the experimental manipulation. While some participants did demonstrate suspicion of the manipulation in this and later gaze-cueing experiments, it was rare for any participant to spontaneously describe the pattern of eye gaze. To offer a qualitative interpretation, many answers were given hesitantly and made it appear that participants were thinking back over the structure of the experiment and inferring from that in order to generate an answer, rather than as a result of their own in-the-moment intuition during gaze-cueing.

Data Analysis

Before data were analysed, participants' responses were filtered to remove all error trials (where participants reported the incorrect answer) and RT outliers − RTs below 250ms (too short to process the stimuli) and above 2,500ms (indicating that participants had not given a response in the allotted time). The number of remaining trials was then compared with the original number of trials to check that all participants retained at least 70% of their total trials and had not scored below 70% total correct on any one condition. Mean RTs and percentage accuracy scores were calculated for each participant for both valid and invalid trials for each block separately. RTs and accuracy rates for each block were compared in separate 2x5 (validity x block) repeated measures ANOVAs.

Average trustworthiness ratings were calculated for each participant both at the beginning (pre) and end (post) of the experiment for both valid and invalid faces, and these scores were analysed in a 2x2 (time x validity) repeated measures ANOVA.

Results and Discussion

Gaze Cueing

The RT results of Experiment 1 are shown in Figure 2a. Over the course of the five blocks, RTs were lower to valid than invalid trials. A 2x5 ANOVA found a main effect of validity ($F(1,23)=47.94$, $p<0.0001$, $\eta^2_G=0.02$) and a main effect of block where responses were faster in later blocks than earlier (using Greenhouse Geisser correction for violation of sphericity assumption: $F(1.92,44.16)=21.51$, $p<0.0001$, $\eta^2_G=0.13$) but no interaction ($F(4,92)=1.14$, $p=0.3448$, $\eta^2_G=0.00$).

[Figure 2 approximately here]

A similar 2x5 ANOVA using accuracy scores (coded as percent correct in each trial type in each of the five blocks; see Table 1) found only a main effect of block ($F(4,92)=3.36$, $p=0.0129$, $\eta^2_G = 0.05$), as participants generally committed more errors at the beginning of the experiment than the end, but there was no overall effect of cueing validity on errors ($F(1,23)=0.15$, $p=0.7002$, $\eta^2_G = 0.00$), and no interaction of validity and block ($F(4,92)=1.46$, $p=0.2204$, $\eta^2_G = 0.01$). These results suggest that the attention cueing effect emerged primarily in RT measures rather than error rates, and that it remained stable over time.

[Table 1 approximately here]

Trustworthiness Ratings

A repeated measures ANOVA with rating time (pre- and post-experiment) and cueing validity of faces (valid and invalid) as within-subjects factors found a significant overall effect of time where ratings were lower after the experiment than before ($F(1,23)=10.49$, $p=0.0036$, $\eta^2_G =0.08$) and one of cueing validity where ratings were lower for invalid faces than valid ($F(1,23)=7.23$, $p=0.0131$, $\eta^2_G =0.13$). There was also a significant interaction between the time of rating and cueing validity ($F(1,23)=7.19$, $p=0.0133$, $\eta^2_G =0.09$). Figure 3a shows how the trustworthiness ratings for each group change over the course of the experiment. It is clear that there are no changes in trust for faces that consistently looked towards targets (Valid: $t(23)=0.07$, $p=0.9455$), whereas there was a significant decline in trustworthiness for faces that consistently looked away from targets (Invalid: $t(23)=-4.19$, $p=0.0003$).

[Figure 3 approximately here]

The results of Experiment 1 demonstrate that the trust effect can be obtained with neutral faces, where previously such effects were not clearly demonstrated (e.g. Bayliss et al., 2009). It is possible that the pre- and post-experiment assessments that allow measures of change in trust for each face, and the use of an unmarked analogue scale, are more sensitive measures than the forced-choice decision used in previous work. Finally, it is noteworthy that there is an asymmetry in the effect in that only invalid faces decline in trustworthiness, while valid faces do not change in trust ratings. Experiment 2 explores whether this holds true when faces express positive emotions.

Experiment 2

This experiment aims to explore how emotion affects this incidental learning of trust. This replicates all details of Experiment 1, but uses smiling rather than neutral face images as the cueing and rating stimuli. Note that Bayliss, et al. (2009) demonstrated significant learning of trust when the faces expressed positive emotions with a smile. However, when the faces expressed a neutral emotion the same pattern of trust was observed, but it was of marginal significance. Experiment 1 has shown that it is possible to detect significant learning of trust when faces are neutral, however the effect was asymmetrical, as invalid faces declined in trust and valid faces did not change. Whether faces expressing positive emotions produce this same pattern is the key question for Experiment 2.

Methods

Participants

24 participants (21 female, mean age 20.46) volunteered for this study in return for course credit or payment.

Stimuli, Design and Procedure

This experiment was identical to Experiment 1 in every way except that the KDEF faces used were frontal-view smiling faces rather than neutral faces both during the gaze-cueing portion of the experiment and at both pre- and post-trustworthiness rating presentations. All other details were identical.

Data Analysis

RT filters and analysis of RTs, error rates and trustworthiness ratings data were identical to those in Experiment 1.

Results and Discussion

Gaze-Cueing

The results of Experiment 2 are shown in Figure 2b. Over the course of the five blocks, RTs were faster to valid than invalid trials. A 2x5 ANOVA found a main effect of validity ($F(1,23)=25.58$, $p<0.0001$, $\eta^2_G=0.02$) and a main effect of block (GG corrected: $F(1.86,42.72)=31.33$, $p<0.0001$, $\eta^2_G=0.25$) but no interaction (GG corrected: $F(4,92)=1.22$, $p=0.3063$, $\eta^2_G=0.00$).

A 2x5 (validity x block) ANOVA on the accuracy rates (see Table 1) found the main effect of block approached but did not reach significance ($F(4,92)=2.07$, $p=0.0912$, $\eta^2_G=0.05$) and there was no effect of validity ($F(1,23)=0.09$, $p=0.7693$, $\eta^2_G=0.00$) or interaction between the two ($F(4,92)=1.27$, $p=0.2874$, $\eta^2_G=0.01$).

Trustworthiness Ratings

The changes in trustworthiness ratings for the faces in Experiment 2 are shown in Figure 3b. A repeated measures ANOVA with rating time (pre- and post-experiment) and cueing validity of faces (valid and invalid) as within-subjects factors found no overall effect of time ($F(1,23)=0.02$, $p=0.9022$, $\eta^2_G=0.00$), but did find a significant effect of cueing validity where valid faces were rated higher than invalid faces ($F(1,23)=5.48$, $p=0.0282$, $\eta^2_G=0.07$), as well as a significant interaction between the time of rating and cueing validity (F(1,23)=12.08, p=0.0020, η^2_G =0.10). As in Experiment 1, there was a decline in trustworthiness ratings for invalid faces that looked away from targets that in this experiment approached significance (Invalid: t(23)=-1.78, p=0.0882), but now we also detect a significant increase in trustworthiness for valid faces that looked towards targets (Valid: t(23)=2.20, p=0.0381).

To explore the contrasts between Experiment 1 and 2 post-hoc tests examining the change in trust for valid faces and invalid faces were examined. Due to violations of the normality assumption, separate Mann-Whitney U tests were performed on the data. When examining the change in trust for valid faces that consistently looked towards targets, increases in trust were greater when the faces were smiling than when neutral (U=181, p=0.0281). In contrast, whether faces were smiling or neutral had no effect on the decline in trust ratings for the invalid faces that looked away from targets (U=282, p=0.9097).[2]

As noted, Bayliss et al. (2009) observed that smiling faces produced more robust trust effects than neutral faces. However, this previous work required forced choice between previously valid and invalid faces. Such forced choice measures cannot identify whether the effects are specific to valid faces increasing in trust, invalid decline in trust, or both. The current approach enables a more detailed analysis where change in trust ratings can identify the specific patterns of trust effects. We can now see that the more robust trust effects are specifically due to increases in trust of valid faces only when they are smiling. Conversely, a decrease in trustworthiness for invalid faces appears to be the stable, key feature of this effect. By comparing trust learning in response to neutral faces with that of smiling faces, we provide the first evidence that this incidental learning is asymmetrical (for a more detailed description of the possible implications and mechanisms underlying this asymmetricality, see the General Discussion).

Based on previous research (Bayliss et al., 2009; Pecchinenda, Pes, Ferlazzo & Zoccolotti, 2008;) we might expect that social expectations driven by emotional expression (i.e. that smiling faces are likely to be helpful) might influence the incidental learning of trustworthiness. That is, valid smiling faces confirm the expectations and so lead to a stronger increase in trust, while invalid smiling faces violate the expectation and so lead to a decrease in trust. In contrast, a neutral stare for social primates can be perceived as a threat and hence the detection of negative events such as potential deception by invalid faces that consistently look away from targets takes precedence and there is little learning of the positive events such as joint attention produced by valid faces.

Experiment 3

Experiments 1 and 2 have confirmed that even when a face is irrelevant and could be ignored while focussing on the main task of peripheral target classification, there is learning of the gaze patterns of another person. That is, an association between face identity and reliability of gaze direction is learned and this is retrieved when re-encountering faces at a later time. Thus, incidental learning of gaze direction subsequently changes how much another person is trusted, and this is affected to some extent by the emotional expression of the person.

The next studies further investigate the properties of this incidental learning process. Experiment 3 explores whether this impression is specific to trust or simply reflects a broader valence impression of the face. Therefore, this experiment is identical to Experiment 1 except that it makes one very minor change: that is, the question that participants are asked was changed from one of “How *trustworthy* do you think this person is?” to “How *likeable* do you think this person is?”

Methods

Participants

26 participants (21 female, mean age 18.99) volunteered for this study in return for course credit or payment. One participant's data were not collected due to a computer malfunction, and 1 participant had to be removed after RT filters were applied, so the final number available for analysis was 24.

Stimuli, Design and Procedure

This experiment was identical to Experiment 1 except that at the beginning and the end of the experiment, participants were asked, "How *likeable* do you think this person is?" rather than "How *trustworthy* do you think this person is?" and all mentions of trustworthiness on consent forms and instructions were replaced with the words likeable or likeability (dependent on context).

Data Analysis

RT filters were applied in the same way as in Experiments 1 and 2, and in this experiment 1 participant had to be removed for retaining less than 70% of their original trials.

Average likeability ratings were calculated for each participant both at the beginning (pre) and end (post) of the experiment for both valid and invalid faces, and these scores were analysed in a 2x2 (time x validity) repeated measures ANOVA.

Results and Discussion

Gaze-Cueing

The results of Experiment 3 are shown in Figure 2c. Over the course of the five blocks, RTs were faster to valid than invalid trials. A 2x5 ANOVA found a main effect of validity ($F(1,23)=14.45$, $p=0.0009$, $\eta^2_G=0.02$) and a main effect of block (using Greenhouse Geisser correction for violation of sphericity assumption: $F(2.18,50.24)=37.73$, $p<0.0001$, $\eta^2_G=0.23$)

but no interaction (GG corrected: $F(2.44,56.13)=1.60$, $p=0.3768$, $\eta^2_G=0.00$).

A 2x5 ANOVA on accuracy rates (see Table 1) found no main effects of block or validity, or any interaction between the two.

Likeability Ratings

The changes in likeability ratings for the faces in Experiment 3 are shown in Figure 3c. A repeated measures ANOVA with rating time (pre- and post-experiment) and cueing validity of faces (valid and invalid) as within-subjects factors found no overall effect of time ($F(1,23)=0.20$, $p=0.6551$, $\eta^2_G=0.00$), or of cueing validity ($F(1,23)=0.04$, $p=0.8371$, $\eta^2_G=0.00$), and no significant interaction between the two ($F(1,23)=2.26$, $p=0.1467$, $\eta^2_G=0.03$).

To explore the contrasts between Experiment 1 and 3 post-hoc tests examining the change in ratings for valid faces and invalid faces were examined. Due to violations of the normality assumption, separate Mann-Whitney U tests were performed on the data. When examining the change in ratings (trust vs. liking) for invalid faces that consistently looked away from targets showed a significantly greater decrease in trust ratings than likeability ratings ($U=166.5$, $p=0.0126$), but no such differences between ratings emerged for valid faces ($U=267$, $p=0.6725$).[3]

The results of Experiment 3 are quite surprising. This experiment was identical to Experiment 1 except for one word change from "trustworthy" to "likeable" in the ratings task. Although the RT gaze cueing effects were the same in Experiments 1 and 3, this gaze cueing had no effect on how likeable a person was perceived to be. It is somewhat surprising that the predictability of gaze-cueing can have such a specific effect, where individuals who consistently look away from targets, a form of deception, are trusted less but not necessarily liked less.

Of course we have to be cautious when interpreting null results, and so in our Supplementary Materials we include an additional version of this experiment (not included here due to it also collecting data from EEG, which makes it difficult to compare with the current package) using different faces and run on a different participant pool (this experiment was conducted at Bangor University rather than the University of York). Critically, this experiment also failed to detect any effect of gaze cueing contingencies on judgements of likeability, whereas significant effects on judgements of trust have been repeatedly observed. Faces used in the supplementary material were pre-selected as appearing high in happiness, which means they expressed slight smiles. Slight smiles are effective at eliciting strong changes in trust (Manssuer et al., 2015a; b), and these are comparable to the effect found in Experiment 2 with full smiles. That we see no learning of liking with these slightly smiling faces leads us to have increased confidence in the lack of incidental learning from eye-gaze on judgments of liking.

Experiment 4

A key feature of the previous research is that trust was influenced by eye-gaze behaviour of another person. Clearly looking towards or away from relevant objects is a means of deceiving another person and initiates joint attention, which recruits reward-related neurocircuitry (Gordon, Eilbott, Feldman, Pelphrey & Vander Wyk, 2013; Schilbach et al., 2010). However, it remains to be seen whether this effect is wholly dependent on this joint attention feature or if similar effects can be induced purely through selective disruptions of visuomotor fluency in the absence of any physical changes to the face. Previous research has shown that perceptual fluency (e.g., Reber et al., 1998) and motor fluency (e.g., Hayes et al., 2008) can influence emotional assessments of stimuli. Can impaired processing of a face with no physical changes, such as eye-gaze shifts, also influence trust judgements? To explore this, Experiment 4 reports a task-switching paradigm designed to match the gaze-cueing paradigm, where participants experience the same disruptions of fluent processing but without any sense of shared attention with the face.

A task-switching paradigm involves participants performing two judgements of a stimulus on different trials. For example, two trials might require the report of the colour of a stimulus, while the next two trials might require the report of the identity of a stimulus. These paired trials and predictable switches between tasks continue throughout an experiment. When the task changes, a visuomotor cost (slower RTs, greater probability of errors) is associated with responses on that switch trial (e.g., Monsell, 2003; Wylie & Allport, 2000; Yeung, Nystrom, Aronson & Cohen, 2006), even when the change sequence is predictable and therefore switches can be anticipated (Kiesel et al., 2010; Rogers & Monsell, 1995). As such, this experiment asks participants to make one of two judgements of face images (colour or sex) where the designated task changes every other trial (a Switch-Repeat alternating runs paradigm).

It is hypothesised that if changing visuomotor fluency evokes affective reactions (see Constable, Bayliss, Tipper & Kritikos, 2013; Hayes et al., 2008) then creating disfluency while processing a particular face identity will reduce trust ratings. That is, throughout the experiment particular face identities are always presented on switch trials where RTs are slowed and errors are more likely, while other face identities are always presented on repeat trials where RTs are fast and accurate. Hence the design matches the gaze-cueing study of Experiment 1 where a particular face identity is always presented on a valid or invalid cueing trial. However, if learning of trust is not simply based on visuomotor fluency, but rather requires specific behaviour associated with deception such as eye-gaze, then simply impairing processing on switch trials in the absence of any physical changes to the face will not change trust ratings.

Methods

Participants

32 participants (29 female, mean age 20.97) volunteered for this study in return for course credit or payment. 8 participants had to be removed after RT filters were applied, and so the final number available for analysis was 24.

Stimuli, Design and Procedure

Stimuli were generated from the same KDEF faces used in Experiment 1. Participants completed the same trustworthiness ratings as in Experiment 1 before and after the experiment, using full colour unaltered images. During the main portion of the experiment, however, the paradigm was changed from gaze-cueing to task-switching, and for this all face images were superimposed with a transparent chromatic hue in Adobe Photoshop CS6 to appear either green or yellow (see Figure 4 for examples). As in Experiments 1-3, each face appeared ten times across the experiment to each participant (twice in each block).

[Figure 4 approximately here]

Participants were told that they would be asked to make one of two judgements about a face image that appeared on the screen; they would either be asked to judge the colour of the image (Colour condition: green or yellow) or to judge the sex of the image (Identity condition: male or female). Participants were told that the task they were to perform would be shown to them as a reminder before each trial, but that the task would change every other trial such that they would perform two Colour trials, then two Identity, and so on. The act of switching between two tasks leads to responses on the first trial of the new task being slower and more error prone; this switching cost to visuomotor fluency was the critical independent variable. As such, half of the identities only appeared immediately after a task-switch, in the first position of the sequence (Switch trial) and half appeared immediately before the switch in the second position (Repeat trial). Identity and trial position was counterbalanced across participants.

During the course of a trial, a condition cue (either 'Colour' or 'Identity' would appear on the screen for 1,000ms, alternating every two trials) to make participants aware of the task they were performing, followed by a 500ms fixation cross. The target image would then appear on the screen for 500ms subtending 23.43˚ visual angle horizontally and 22.62˚ vertically, followed by a blank screen for 1,000ms. Participants could respond at any point in this 1,500ms window, and any response after this time window was classed as incorrect. Participant responses were the keyboard buttons Z and M, each of which corresponded to a different answer in the two tasks (i.e. Z, male and green; M, female and yellow − counterbalanced across participants).

The same RT filters were applied to the data as in Experiments 1, 2 and 3 with the difference that the 2,500ms upper limit was shortened to 1,500ms to reflect the timings of the experiment. Incorrect responses and responses faster than 250ms were removed from the data and the participants' accuracy and number of trials were considered to see if they retained more than 70% of their original number of trials. This paradigm proved to be more difficult for participants than gaze-cueing, as 8 participants committed too many errors to be suitable for inclusion.

RT and accuracy rates were calculated for each participant for both Switch and Repeat trials in each of the five blocks, and these were compared in separate 2x5 repeated measures ANOVAs. Trustworthiness ratings were calculated in the same way as in Experiment 1 with Switch/Repeat replacing invalid/valid as the independent measure and these were analysed in a 2 (pre/post-experiment) x 2 (Switch/Repeat) repeated measures ANOVA.

Results and Discussion

Task-Switching

The results of Experiment 4 are shown in Figure 2d. Over the course of the five blocks, RTs were faster to Repeat than Switch trials. A 2x5 ANOVA found a main effect of trial ($F(1,23)=27.03$, $p<0.0001$, $\eta^2_G=0.05$) and a main effect of block (GG corrected: $F(2.64,57.99=5.11$, $p=0.0010$, $\eta^2_G=0.04$) but no interaction ($F(4,92)=1.74$, $p=0.1480$, $\eta^2_G=0.00$).

A 2x5 ANOVA on accuracy rates (see Table 1) found a main effect of trial ($F(1,23)=17.58$, $p=0.0003$, $\eta^2_G = 0.05$) with more errors on Switch trials that Repeat trials and a main effect of block (GG corrected: $F(2.58,59.39)=9.33$, $p<0.0001$, $\eta^2_G = 0.22$) but no interaction (F(4,92)=1.39, p=0.2430, $\eta^2_G=0.00$).

Trustworthiness Ratings

The changes in trustworthiness ratings for the faces in Experiment 4 are shown in Figure 3d. A repeated measures ANOVA with rating time (pre- and post-experiment) and position of face in the task sequence (Switch and Repeat) as within-subjects factors found a significant effect of time with more positive trustworthiness ratings after the experiment than before ($F(1,23)=5.02$, $p=0.0351$, $\eta^2_G=0.13$), but none of face position ($F(1,23)=0.42$, $p=0.5213$, $\eta^2_G=0.00$), and no interaction between the two ($F(1,23)=1.31$, $p=0.2636$, $\eta^2_G=0.00$).

These results suggest that in the absence of any physical changes to the face that led to the initiation of joint attention with participants, the change in trustworthiness judgements was not replicated. This suggests that the trust effect cannot be explained purely by disruptions of visuomotor fluency.

Experiment 5

Experiment 4 shows that disruptions to visuomotor fluency using a task-switching paradigm are not sufficient to change judgements of trustworthiness. However, there were a number of methodological changes between Experiment 1 and Experiment 4, not least of which was that the latter made the faces targets of participants' judgements, rather than distractors (we thank a reviewer for highlighting this). Literature on distractor devaluation suggests that to-be-ignored information often shows a devaluation that to-be-attended information does not (see Raymond, 2009). It could be that we do not see any learning of trust (particularly the characteristic decrease for invalid faces evident in Experiments 1 and 2) in Experiment 4 because it is more difficult to devalue targets than distractors. If that is the case, a similar task-switching experiment to Experiment 4, but where faces are presented as background distractors, as in Experiments 1-3, might result in similar incidental learning patterns to those in Experiments 1-3. Experiment 5 explores this and adapts the task-switching paradigm of Experiment 4 to more closely match the object-categorisation task that participants completed in Experiments 1-3.

Methods

Participants

28 participants (20 female, mean age 19.07) volunteered for this study in return for a mixture of course credit and payment. 4 participants had to be removed after RT filters were applied, and so the final number available for analysis was 24.

Stimuli, Design and Procedure

[Figure 5 approximately here]

This experiment closely matched the gaze-cueing experiment used in Experiment 1, but the faces no longer shifted their gaze. Instead, participants were told that they would be making one of two possible judgements on a given trial; the first was object TYPE, where they would categorise the object as either a kitchen or garage item (as in other experiments), while the second was object COLOUR, where they would judge whether the object was blue or yellow.

Changes to the stimuli from previous experiments were the introduction of yellow-coloured objects (the same objects as used in previous experiments but digitally manipulated to appear yellow instead of blue) and the fact that when faces appeared in the centre of the screen we used unaltered, neutral images rather than those digitally manipulated to shift their gaze. We also introduced a task cue before each trial, to remind participants of whether they were supposed to judge the object's TYPE (kitchen/garage) or COLOUR (blue/yellow; see Figure 5).

The task that participants completed altered every other trial in a Switch/Repeat task-switching procedure, as in Experiment 4, with each face presented ten times over the course of the experiment (twice in each of the five blocks), and a feedback tone was presented for incorrect responses. Participants were once again instructed to ignore the faces as irrelevant. Participants completed trustworthiness ratings both at the beginning and the end of the experiment.

Results and Discussion

Task-Switching

The results of Experiment 5 are shown in Figure 2e. Over the course of the five blocks, RTs were faster to Repeat than Switch trials. A 2x5 ANOVA found a main effect of trial

(F(1,23)=37.92, p<0.0001, η^2_G =0.06) but no main effect of block (GG corrected: F(2.53,58.10=0.51, p=0.7270, η^2_G =0.01) and no interaction (GG corrected: F(2.90,66.77)=0.25, p=0.9079, η^2_G =0.00).

A 2x5 ANOVA on accuracy rates (see Table 1) found a main effect of block (GG corrected: F(3.12,71.67)=3.34, p=0.0225, η^2_G =0.06), as well as a main effect of trial (F(1,23)=49.56, p<0.0001, η^2_G =0.07) but no interaction between the two (F(4,92)=0.79, p=0.5330, η^2_G =0.01).

Trustworthiness Ratings

The changes in trustworthiness ratings for the faces in Experiment 4 are shown in Figure 3e. A repeated measures ANOVA with rating time (pre- and post-experiment) and position of face in the task sequence (Switch and Repeat) as within-subjects factors found an effect of face position (F(1,23)= 5.74, p=0.0251, η^2_G =0.07), but found no significant effect of time in this experiment (F(1,23)=0.69, p=0.4132, η^2_G =0.01) and no interaction between the two (F(1,23)=0.18, p=0.6718, η^2_G =0.00).

To explore the contrasts between Experiment 1 and 5 we examined the changes in ratings for valid faces and invalid faces separately. In this contrast the assumption of normality was not violated, and so independent t-tests were used. When examining the change in trustworthiness ratings for low-fluency faces (Switch and invalid faces) there was a significantly greater decrease in trust ratings during gaze cueing than task switching (t(46)=-3.70, p=0.0006), but no such difference emerged for valid faces (t(46)=-0.50, p=0.6218).[4]

Although the results of Experiment 5 do show a significant effect of face position, this appears to be due to chance differences in the pre-ratings – there is no logical reason to suppose that visuomotor fluency could have an effect before participants encounter it, and so these differences must be due to random chance. The fact that they do not change over the course of the experiment, as evidenced by the lack of interaction and the remarkably flat profile of changes, is evidence that this incidental learning cannot be explained in terms of visuomotor fluency – even when accounting for whether the faces are targets or distractors.

However, the lack of any effect of time in this experiment suggests that the overall increase in trustworthiness seen in Experiment 4 may indeed have been due to the faces' status as target rather than distractor stimuli, and reflect a familiarity effect that is not evident here.

General Discussion

Detecting and learning about subtle cues to trustworthiness is of critical importance during social interactions. One such cue is the eye-gaze pattern of another individual, whether they are reliable and look towards relevant objects in a scene or deceive by looking away from objects. As shown in the RT results of the first three experiments, gaze-cues are encoded rapidly and automatically and hence are effective ways of misdirecting others, as shifts of attention are difficult to inhibit. A series of experiments have further investigated the boundary conditions of the learning of trust from patterns of eye gaze.

Experiment 1 demonstrated that learning of trust is possible even when faces express neutral emotions. Previous work highlighted the role of emotion in these gaze-trust effects. Bayliss et al., (2009) demonstrated significant trust learning effects when faces smiled, no effects when they frowned, and marginal effects when the faces were neutral.

Experiment 1 has revealed that significant trust effects can be obtained with neutral faces. It is unclear whether the failure to detect effects in the Bayliss et al. study was a Type I error, or whether the changes to the procedure were of critical importance. In the previous work a two alternative forced choice (2AFC) task was employed where pairs of faces that had consistently looked towards targets (valid) or had looked away from targets (invalid), were presented and participants selected the one who they felt was more trustworthy. In contrast, the current study requires assessment of trust for each individual face and it measures changes in trust ratings from the start to the end of the experiment.

We feel this new approach is a more sensitive and robust means of measuring trust. Furthermore it provides important information concerning where the effect may lie. That is, 2AFC can only reveal that faces that previously looked towards targets tend to be selected as more trustworthy, not whether valid faces are trusted more, invalid faces trusted less, or both. The results of Experiment 1 suggest an asymmetry, where the effect is only observed in the decline in trust of invalid faces that looked away from targets, whereas there is no change in trust rating for the valid faces that always looked towards targets.

That this effect initially manifests as a decrease in trust only for invalid faces may relate to the finding by both Bayliss and Tipper (2006) and Bayliss et al., (2009) of a memory bias for invalid faces. That is, the false belief that invalid faces appeared more frequently than valid faces during the course of the gaze-cueing experiment. This builds on other work showing memory advantages for cheaters (Bell et al., 2012; Buchner, Bell, Mehl & Musch, 2009), and suggests that the motivation behind this trust effect is to remember those faces that frequently present challenges to visuomotor fluency – therefore there is less change in trust for valid faces simply because participants are not motivated to remember those identities as clearly.

The second aim of this paper was to explore the role that emotion plays at initial encoding. Experiment 2 demonstrates that there appears to be a change in the pattern of trust learning when the faces smile; that is, in contrast to Experiment 1 where effects were only detected in a decline in trust for invalid faces. When the faces smile a bi-directional effect is observed, where invalid faces again show a decrease in trust, while valid faces now produce a significant increase in trust. This latter bi-directional effect with smiling faces has also been demonstrated by Manssuer et al. (2015a,b).

There are multiple potential explanations for the difference in the pattern of results between Experiments 1 and 2 that future research should investigate. For example, the default learning mechanism might be to detect deception. Certainly in terms of memory for faces, this is better for faces that deceive (Bayliss et al., 2009; Bayliss & Tipper, 2006; Bell et al., 2012; Buchner et al., 2009), hence learning of trust is only evident in invalid faces that deceive and look away from targets. In contrast, when the faces all express positive emotion, this combines with the positive signal of joint attention evoked by valid cueing faces, hence increasing trust of these faces. Alternatively, the positive social context motivates participants to remember the faces in the experiment. As invalid faces are apparently remembered well regardless of emotion (given the similar trust change profile for invalid faces across Experiments 1 and 2), this seems to primarily affect valid faces.

The aim of the second half of this paper was to explore some of the boundary conditions of this effect. Experiment 3 replaced the question of trustworthiness that participants were asked with a question of likeability; simply by changing a single word in the design the effect was abolished. The lack of an effect when judging liking is somewhat counterintuitive. Therefore it was of value to report a further experiment in the supplementary materials. This study had a number of procedural differences to Experiment

3, the most important of which was that the faces expressed positive emotions (slight smiles). This situation is closer to Experiment 2 where positive smiles were observed, and again failed to show any effects when assessing liking of another person. Hence we doubt this is a Type 2 error.

This lack of effect with liking judgements suggests that the gaze-contingent trust effect is highly specific to trust − a fact that makes sense if one considers that trust as a trait judgement serves much more heavily as a predictive model of behaviour than does liking: we decide how much to trust someone based on how we expect them to behave, whereas liking is a more subjective and affective judgement, and one less based in statistical contingencies. For example, incidental learning of gaze contingencies will influence economic decisions to invest in another person (e.g., Rogers et al., 2014). It is possible that if we manipulated participants' beliefs about intentions or sense of competition that an effect of liking would emerge, but at its basic level this effect appears to be specific to monitoring the trustworthiness of interactants. To our knowledge this provides the first evidence of a functional distinction between trust and liking, and directions for future research may examine other ways in which these two differ, and the possible mechanisms underlying each of them.

It is now well established that the eye movements of another person automatically shift attention and whether they consistently look towards or away from objects mediates incidental learning of trust. The shifts of attention of another person certainly can be used to deceive, and hence it might be predicted that if there are no such behaviours in a face, then learning of trust does not take place, even though particular face identities are associated with different levels of visuomotor fluency. Therefore Experiments 4 and 5 examined whether learning of trust could be generated in the absence of any physical changes to the faces through a task switching procedure. We found that in the absence of any physical changes, disruptions to participants' sense of visuomotor fluency were not sufficient to generate changes in trustworthiness, despite the RT costs associated with task-switching being comparable to those associated with gaze-cueing (see Figure 2). This finding also held true regardless of whether the faces were targets (Experiment 4) or distractors (Experiment 5).

This contrasts with previous work that has shown that perceptual fluency does increase liking of objects (e.g. Reber et al., 1998; Zajonc, 1968). It is also worth noting that this cannot be explained by the faces being more resilient to devaluation in Experiment 4, as Experiment 5 uses the same faces as distractors that appear before the target object. Taken together, these two results provide strong evidence against disruptions to visuomotor fluency being sufficient for incidental learning of trust.

There is also previous literature that does report that processing fluency can affect judgements of trust. For example, Winkielman, Olszanowski and Gola (2015) found that increasing the disfluency associated with certain faces in an emotion categorisation procedure led to decreased ratings of trust in later judgements, and that the effect of this disfluency was unrelated to face valence. However, it is important to note that our experiments examine learning in the absence of physical cues to trustworthiness (such as changes in expression) and as such these results are not necessarily inconsistent with this previous literature. Our interpretation fits with our earlier point that this incidentally learned trust reflects a sense of reliability (or unreliability) rather than variations along a dimension of warmth.

Learning of trust from patterns of eye-gaze is probably effective because joint attention can be positively reinforcing (Schilbach et al., 2010) and because gaze direction can be used by primates and humans to misdirect the attention of others (e.g. Klein, Shepherd & Platt, 2009). Hence the invalid gaze-cue when the face looks away from the highly salient target will be perceived as an act of deception. In contrast, the static faces in the task switching procedure of Experiments 4 and 5 do not provide such socially relevant information (see also Manssuer et al, 2015a, b). However, we note that some other procedures might be able to influence trust judgements. For example, Fenske, Raymond, Kessler, Westoby & Tipper (2005) demonstrated that response inhibition associated with a face reduced ratings of trust. Hence we can conclude that visuomotor fluency is not sufficient to produce changes in trust, but we are yet to identify other processes that might influence trust ratings.

A final point to be addressed concerns the question of how much these results may be due to demand characteristics. After concluding the experiment several participants were able to deduce the aims when prompted, but few had spontaneous or confident answers. It is likely that the mere act of enquiring about their beliefs may have triggered a reassessment of the experiment, developing their answer as a sensible explanation without previously considering it. This seems particularly likely given that Experiment 3 was as clear to interpret as Experiments 1 and 2, and yet showed no evidence of incidental learning in the likeability ratings. However, we are not making strong claims concerning conscious awareness of learning. Rather, our learning task is incidental in that the faces were irrelelvant to a participant's somewhat demanding main goal of identifying peripheral objects.

In sum, the experiments presented here replicate and support previous findings showing that participants learn incidentally presented identity-gaze contingencies, and use them to inform subsequent trust judgements. We found that with faces expressing a neutral emotion, this trust effect was driven primarily by a decrease in trust for invalid-cueing faces, which perhaps reflects enhanced memory for individuals who deceive. In contrast, when faces express positive emotions learning of trust for valid faces emerges. We also found that the effect is specific to judgements of trust and does not generalise to other judgements, such as liking. Finally, reducing visuomotor fluency is not sufficient to influence ratings of trust; physical changes to the face such as gaze shifts that encourage or disrupt feelings of joint attention and deception play a role. Taken together, these experiments have further specified the properties of the incidental trust learning system.

References

Bayliss, A. P., Griffiths, D. & Tipper, S. P. (2009). Predictive gaze cues affect face evaluations: The effect of facial emotion. *Eur J Cogn Psychol*, *21* (7), 1072–1084.

Bayliss, A. P., Paul, M. A., Cannon, P. R. & Tipper, S. P. (2006). Gaze cuing and affective judgements of objects: I like what you look at. *Psychonomic Bulletin and Review*, *13* (6), 1061–1066.

Bayliss, A. P., Griffiths, D. & Tipper, S. P. (2009). Predictive gaze cues affect face evaluations: The effect of facial emotion. *Eur J Cogn Psychol, 21*(7), 1072-1084.

Bayliss, A. P. & Tipper, S. P. (2006). Predictive gaze cues and personality judgements: Should eye trust you? *Psychol Sci*, *17* (6), 514–520.

Bell, R., Buchner, A., Erdfelder, E., Giang, T., Schain, C. & Riether, N. (2012). How specific is source memory for faces of cheaters? Evidence for categorical emotional tagging. *J Exp Psychol Learn Mem Cogn*, *38* (2), 457–472.

Buchner, A., Bell, R., Mehl, B. & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behaviour*, *30* (3), 212–224.

Capozzi, F., Bayliss, A. P., Elena, M. R. & Becchio, C. (2015). One is not enough: Group size modulates social gaze-induced object desirability effects. *Psychon Bull Rev*, *22*(3), 850-855.

Constable, M.D., Bayliss, A. P., Tipper, S. P. & Kritikos, A. (2013). Cognitive fluency as an alternative route to preference formation. *Conscious Cogn*, *22* (1), 47–52.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E. & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6* (5), 509–

540.

Fenske, M.J., Raymond, J.E., Kessler, K., Westoby, N. & Tipper, S.P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychological Science,* 16(10), 753-758.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, *11*(2), 77-83.

Friesen, C. K. & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, *5* (3), 490–495.

Frischen, A. & Tipper, S. P. (2004). Orienting Attention Via Observed Gaze Shift Evokes Longer Term Inhibitory Effects: Implication for Social Interactions, Attention, and Memory. *Journal of Experimental Psychology: General*, *133* (4), 516–533.

Gordon, I., Eilbott, J. A., Feldman, R., Pelphrey, K. A. & Vander Wyk, B. C. (2013). Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses. *Soc Neurosci*, *8* (6), 544–554.

Halberstadt, J., Winkielman, P., Niedenthal, P. M. & Dalle, N. (2009). Emotional conception: how embodied emotion concepts guide perception and facial action. *Psychol Sci*, *20* (10), 1254–1261.

Hayes, A. E., Paul, M. A., Beuger, B. & Tipper, S. P. (2008). Self produced and observed actions influence emotion: the roles of action fluency and eye gaze. *Psychol Res*, *72* (4), 461–472.

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and Dynamic Facial Cues Differentially Affect the Consistency of Social Evaluations. *Personality and Social*

Psychology Bulletin, *41*(8), 1123-1134.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M. & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136* (5), 849-874.

Klein, J. T., Shepherd, S. V. & Platt, M. L. (2009). Social attention and the brain. *Current biology: CB*, *19* (20), R958–R962.

Lundqvist, D., Flykt, A. & Öhman, A. (1998). The Karolinska directed emotional faces. *Stockholm, Sweden: Karolinska Institute*.

Manssuer, L. (2015). Psychophysiological evidence of a role for emotion in the learning of trustworthiness from eye-gaze cues. (Unpublished PhD thesis). Bangor University: Wales.

Manssuer, L., Pawling, R., Hayes, A. E. & Tipper, S. P. (2015a). The Role of Emotion in Learning Judgements of Trustworthiness from Eye Gaze: Evidence from Facial Electromyography. *Cognitive Neuroscience.*

Manssuer, L. R., Roberts, M. V. & Tipper, S. P. (2015b). The late positive potential indexes a role for emotion during learning of trust from eye-gaze cues. *Social Neuroscience, 10*(6), 635-650.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7* (3), 134–140.

Oosterhof, N. N. & Todorov, A. (2008). The functional basis of face evaluation. *PNAS*, *105* (32), 11087-11092.

Oosterhof, N. N. & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9* (1), 128–133.

Pecchinenda, A., Pes, M., Ferlazzo, F. & Zoccolotti, P. (2008). The combined effect of gaze direction and facial expression on cueing spatial attention. *Emotion*, *8* (5), 628–634.

Raymond, J. (2009). Interactions of attention, emotion and motivation. *Progress in brain research*, *176*, 293-308.

Reber, R., Winkielman, P. & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgements. *Psychol Sci*, *9* (1), 45–48.

Rogers, R. D. & Monsell, S. (1995). Costs of a Predictable Switch between Simple Cognitive Tasks. *Journal of Experimental Psychology-General*, *124* (2), 207–231.

Rogers, R. D., Bayliss, A. P., Szepietowska, A., Dale, L., Reeder, L., Pizzamiglio, G., . . . Tipper, S. P. (2014). I Want to Help You, But I Am Not Sure Why: Gaze-Cuing Induces Altruistic Giving. *Journal of Experimental Psychology: General*, 143,(2), 763-777.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds Made for Sharing: Initiating Joint Attention Recruits Reward-related Neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715.

Ulloa, J. L., Marchetti, C., Taffou, M. & George, N. (2015). Only your eyes tell me what you like: Exploring the liking effect induced by other's gaze. *Cogn Emot*, *29*(3), 460-470.

Wang, Y., & Hamilton, A. F. D. C. (2014). Why does gaze enhance mimicry? Placing gaze-mimicry effects in relation to other gaze phenomena. *The Quarterly Journal of Experimental Psychology*, *67*(4), 747-762.

Winkielman, P., Olszanowski, M., & Gola, M. (2015). Faces in-between: Evaluations reflect the interplay of facial features and task-dependent fluency. *Emotion*, *15*(2), 232.

Wylie, G. & Allport, A. (2000). Task switching and the measurement of ”switch costs”.

Psychological Research-Psychologische Forschung, *63* (3-4), 212–233.

Yeung, N., Nystrom, L. E., Aronson, J. A. & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *J Neurosci*, *26* (5), 1429–1438.

Zajonc, R. B. (1968). Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology Monograph Supplement*, *9* (2), 1–27.

Footnotes

1. Although means and standard deviations were not retrieved from Oosterhof & Todorov's (2008) material, we can validate our assumption that these groups of faces were close to neutral by examining the pre-ratings assigned to them in the five experiments presented here. As the pre-ratings occurred before any participants had a chance to experience the faces within an experimental context, any differences could only be explained by physical cues to trustworthiness, and the combined power of these five experiments would be sufficient to detect this. We explored this and found that the pre-ratings for faces in one group (M=-2.71, s.d.=13.04) did not significantly differ from the other (M=0.85, s.d.=6.82; t(14)=-0.68, p=0.5055).
2. For the sake of completeness we also compared the results of Experiments 1 and 2 in a mixed 2 (cueing validity; within) x2 (experiment; between) ANOVA, with change in trustworthiness (calculated as pre-experiment ratings subtracted from post-experiment ratings) as the dependent variable, and with planned contrasts between valid and invalid faces across experiments. We found a significant overall effect of validity ($F(1,46)=19.21$, $p=0.0001$, $\eta^2_G=0.14$) but none of experiment ($F(1,46)=1.56$, $p=0.2182$, $\eta^2_G=0.02$) and no interaction between the two ($F(1,46)=0.77$, $p=0.3860$, $\eta^2_G=0.01$). Planned comparisons found that the difference between neutral and smiling valid faces approached significance ($p=0.0791$), but there was no such difference between neutral and smiling invalid faces ($p=0.6235$).
3. Comparing Experiments 1 and 3 with a 2x2 mixed ANOVA found a significant overall effect of validity ($F(1,46)=9.44$, $p=0.0040$, $\eta^2_G=0.10$) as well as an effect of experiment ($F(1,46)=4.94$, $p=0.0312$, $\eta^2_G=0.05$) but no interaction between the two

(F(1,46)=2.35, p=0.1322, η^2_G =0.03). Planned comparisons showed the difference between trustworthiness and likeability ratings was significant for invalid faces (p<0.0001) but not for invalid faces (p=0.5020).

4. Comparing Experiments 1 and 5 in a 2x2 mixed ANOVA found a main effect of fluency (high – valid or repeat; low – invalid or switch; F(1,46)=6.72, p=0.0127, η^2_G =0.07) as well as an effect of experiment (F(1,46)=8.02, p=0.0068, η^2_G =0.08) and a significant interaction between the two (F(1,46)=4.88, p=0.0322, η^2_G =0.05). Planned comparisons showed that high fluency (valid and Repeat) faces did not differ significantly between the gaze-cueing and task-switching paradigms (p=0.5090), but there was a highly significant difference between low-fluency (invalid and Switch) faces (p<0.0001). We do not include a similar analysis for Experiment 4 as this was much more methodologically distinct from Experiment 1 and so such comparison would be difficult to interpret.

Table 1. Accuracy rates (% correct) in response to valid/Repeat and invalid/Switch faces across five blocks in Experiments 1-4.

Experiment	*Trial*	Block 1	Block 2	Block 3	Block 4	Block 5
1	*Valid*	84.90	86.20	90.63	91.67	93.50
	Invalid	85.42	88.54	89.84	90.89	90.10
2	*Valid*	90.89	89.06	95.31	90.10	95.31
	Invalid	90.36	90.10	94.53	93.23	93.75
3	*Valid*	91.93	89.84	95.83	92.71	92.19
	Invalid	88.02	88.80	93.23	92.45	95.05
4	*Repeat*	71.09	86.46	88.02	86.46	90.63
	Switch	67.45	83.33	83.85	81.51	89.84
5	*Repeat*	93.49	97.14	96.61	97.92	97.14
	Switch	91.15	91.67	93.75	95.31	94.53

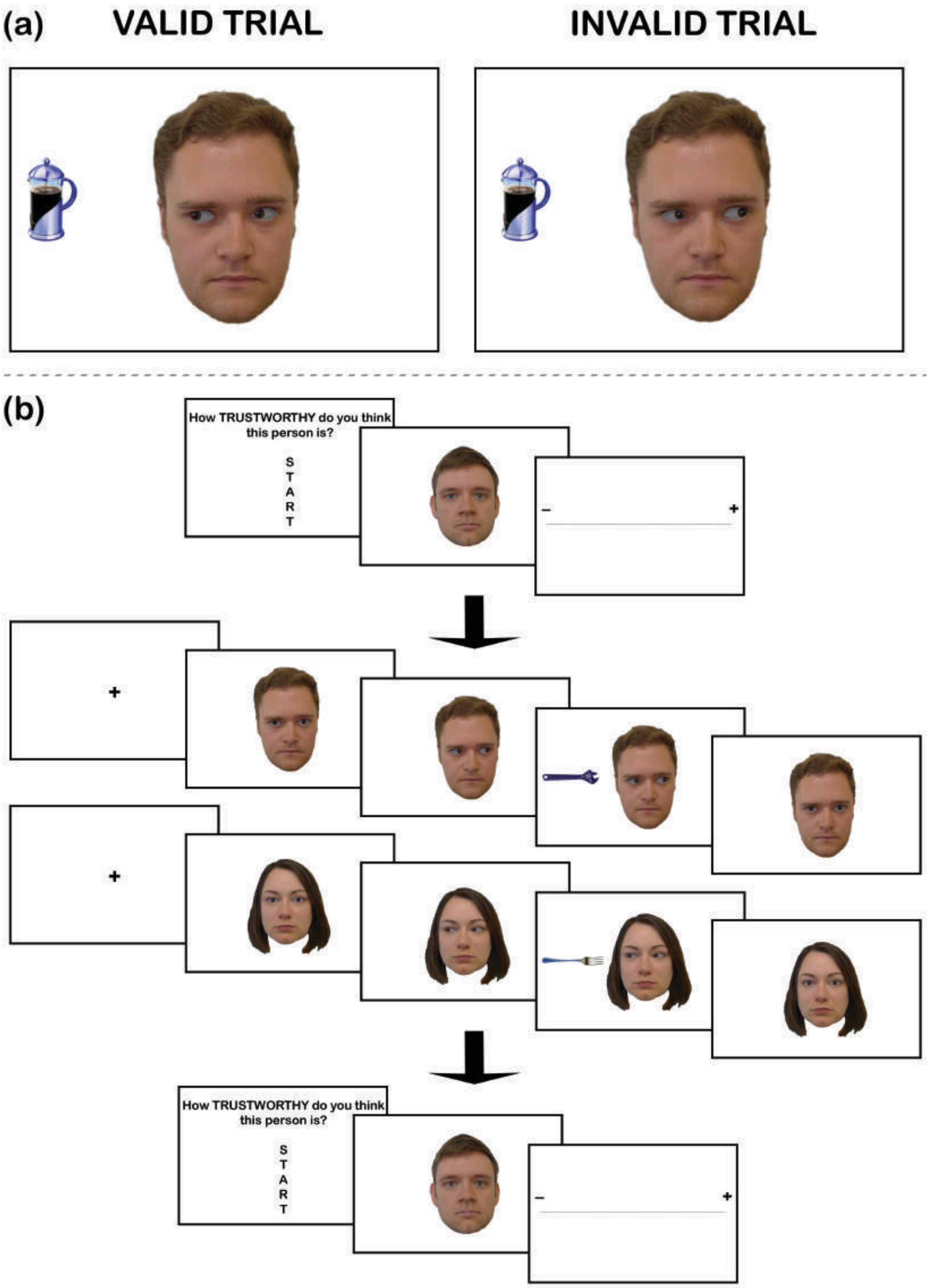

Figure 1. Outline of gaze-cueing procedure used in Experiments 1-3. (a) Examples of a single face on a valid (left) and invalid (right) trial. A participant would see this face in only one of the two conditions. (b) The trial sequence of the whole experiment. Participants made

ratings of the faces (trustworthiness in Experiments 1 and 2, shown, likeability ratings in Experiment 3) at the beginning (top) and end (bottom) of the experiment, and in the main body participants categorized the kitchen and garage objects with key-press responses while ignoring the faces.

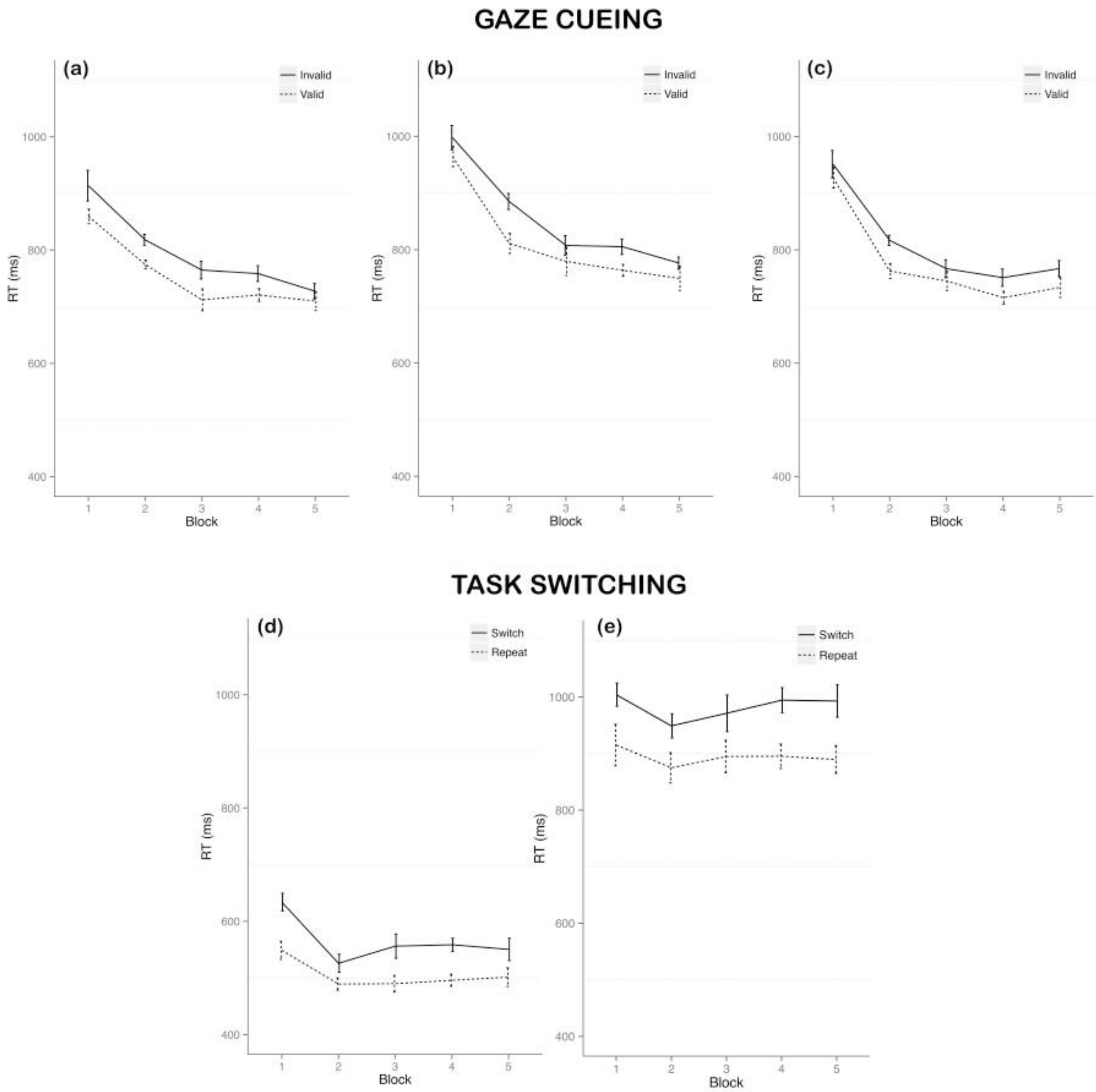

Figure 2. Line graphs tracking changes in reaction times (RT; ms) across five blocks in response to valid (dotted line) and invalid (solid line) trials for (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, and repeat (dotted line) verses switch (solid line) trials in (d) Experiment 4 and (e) Experiment 5. Error bars show standard error.

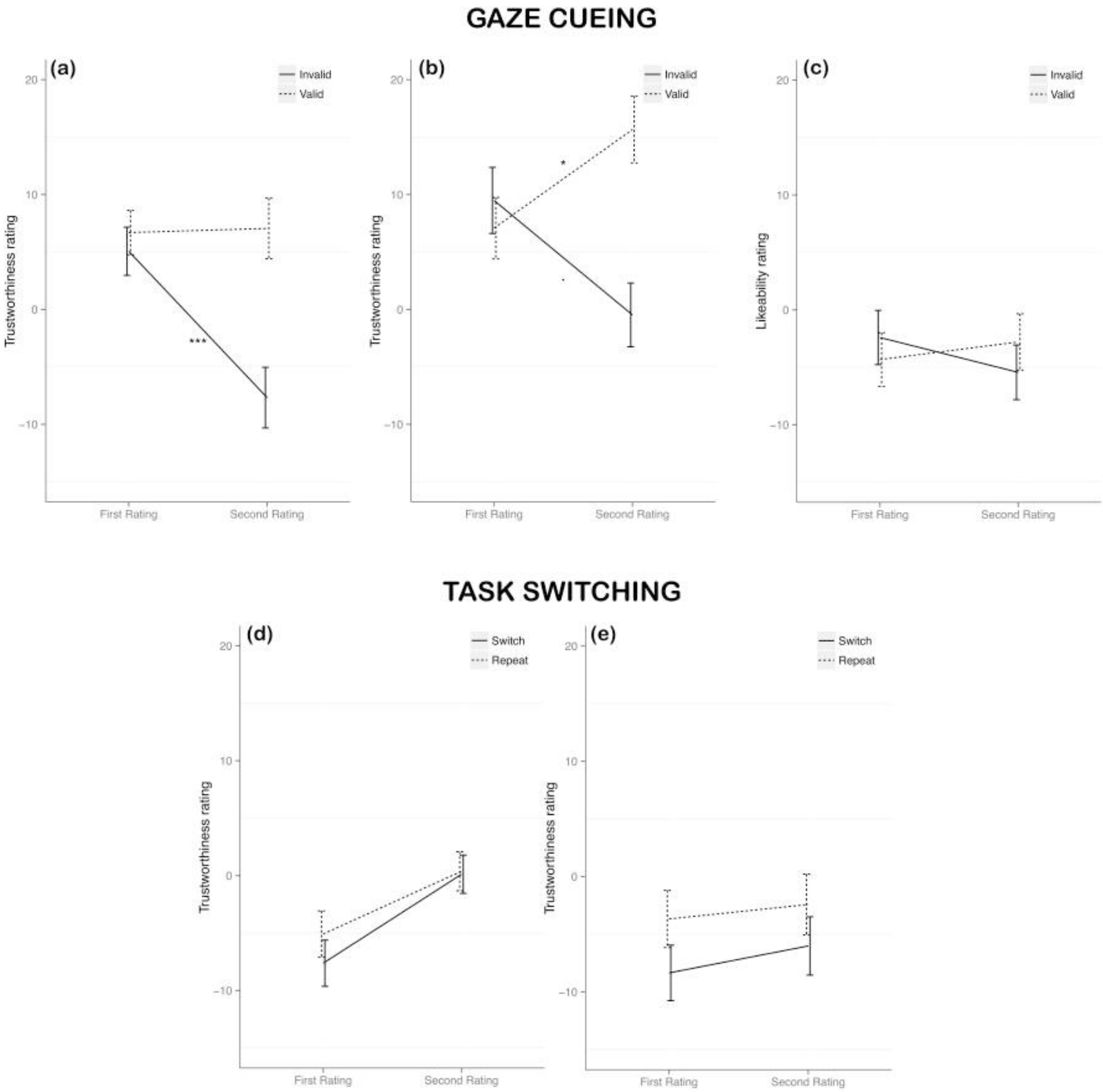

Figure 3. Line graphs tracking changes in face ratings from the beginning to the end of the experimental session for valid (dotted line) and invalid (solid line) trials for (a) Experiment 1: neutral faces, trustworthiness; (b) Experiment 2: smiling faces, trustworthiness; (c) Experiment 3: neutral faces, likeability; (d) for faces used in repeat (dotted line) and switch trials (solid line) in Experiment 4: neutral faces as targets with no gaze movements, task-switching; and (e) for faces used as task-switching distractors with no gaze movements in Experiment 5. Error bars show standard error. †p<.1; *p<.05; ***p<.001

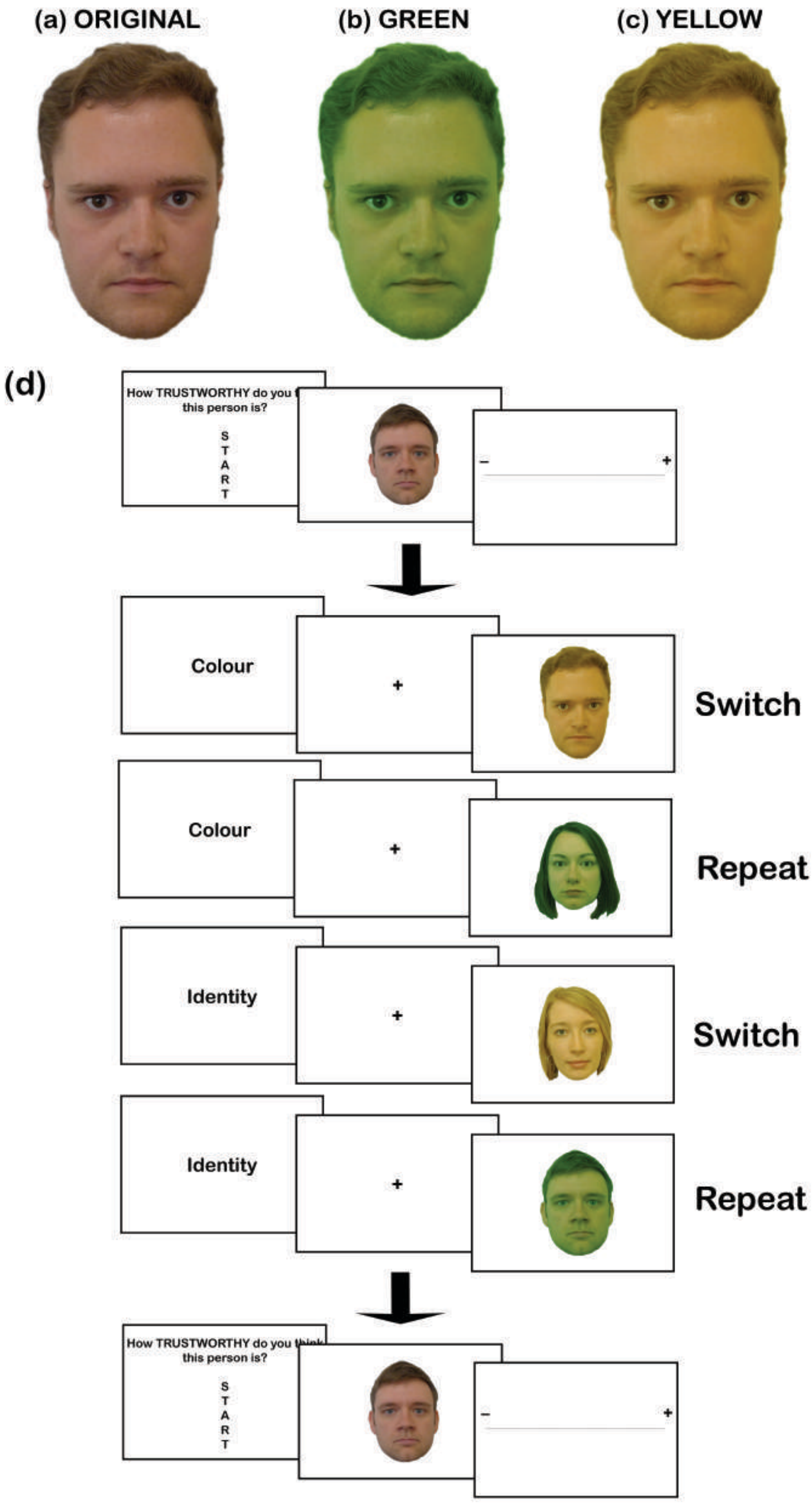

Figure 4. Examples of the coloured stimuli used in the task-switching experiment. (a) The original uncoloured images were used during trustworthiness ratings, while (b) the green

and (c) yellow were used in the task-switching portion. (d) Trial sequence. Participants reported whether the face was coloured in green or yellow or if the face was male or female, depending on a prompt before each trial.

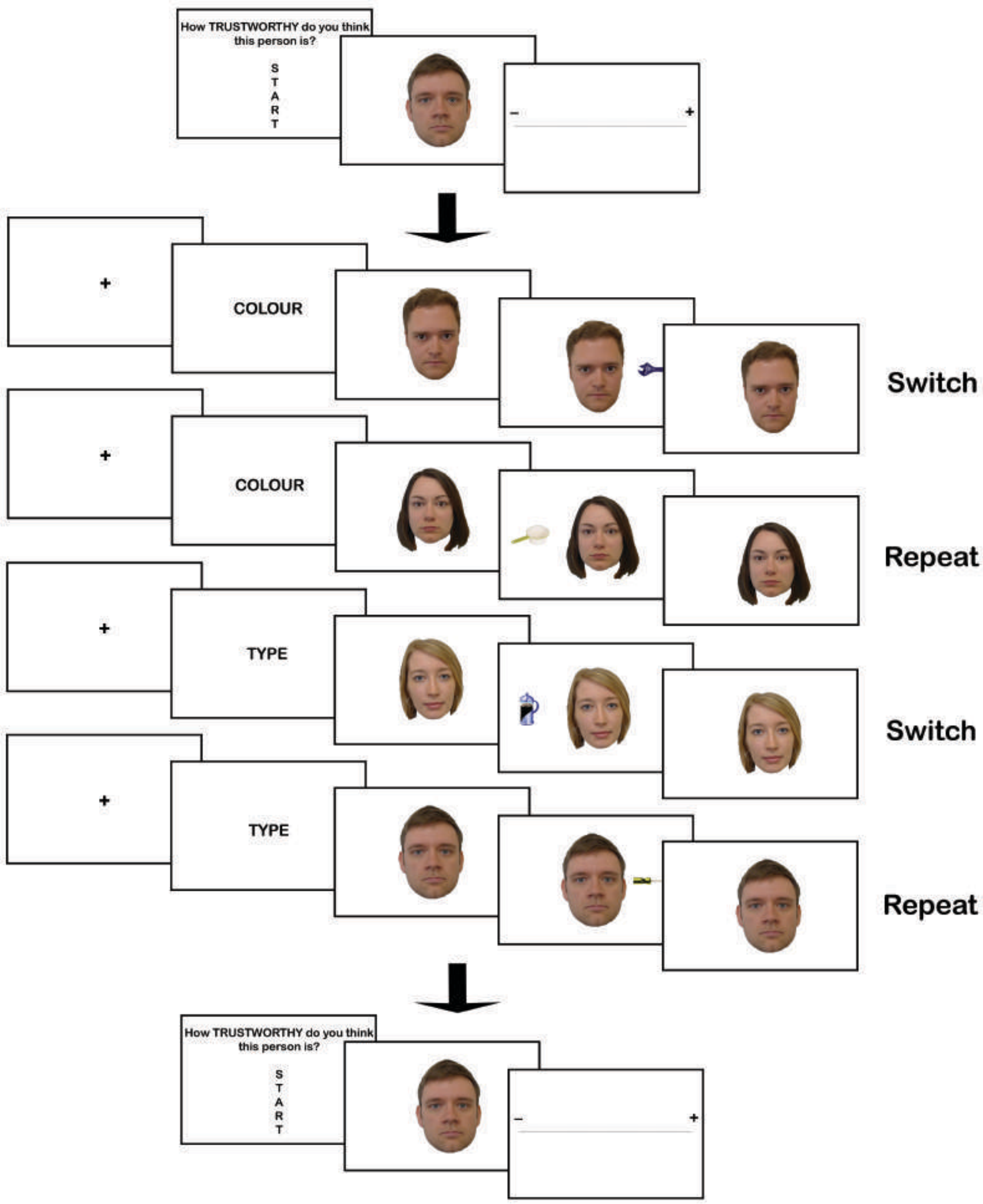

Figure 5. Example trials from Experiment 5. Points to note include the task prompt at the start of each trial, as in Experiment 4, which was then followed by the (unmanipulated) face image, then the object, as in Experiments 1-3.